

Direct semi-parametric estimation of the state price density implied in option prices

Gianluca Frasso ^{*} & Paul H.C. Eilers [†]

ABSTRACT

We present a direct semi-parametric approach for the estimation of the State Price Density (SPD) implied in quoted option prices. We treat the observed prices as expected values of possible pay-offs at maturity weighted by the unknown probability density function. We model the logarithm of the latter as a smooth function while matching the expected values of the potential pay-offs with the observed prices. This leads to a special case of the penalized composite link model. Our estimates do not rely on any parametric assumption on the underlying asset price dynamics and are consistent with no-arbitrage conditions.

1 Introduction

Under equilibrium conditions, the value of an option contract is equal to the discounted expected value of its future net returns. The expectation is taken with respect to a probability density function known as state price density (SPD), risk-neutral density (Cox and Ross, 1976) or equivalent martingale measure (Harrison and Kreps, 1979). The state price density is a fundamental tool for pricing and hedging. It simplifies the evaluation of illiquid instruments or of contracts which values are difficult to workout analytically (e.g. exotic options), and contains important information on market behavior, expectations and preferences. The SPD is not directly observed. However, under restrictive assumptions on the price dynamics of the option's underlying asset, its functional form can be defined analytically. For example,

^{*}Institut des Sciences Humaines et Sociales, Méthodes Quantitatives en Sciences Sociales, Université de Liège, Belgium.

Email: Gianluca.frasso@ulg.ac.be

[†]Department of Biostatistics, Erasmus Medical Centre, Rotterdam, The Netherlands.

Email: p.eilers@erasmusmc.nl

the well known Black and Scholes (1973, shortly B&S in what follows) pricing model assumes that the underlying asset prices follow a geometric Brownian motion process from which a log-normal risk-neutral density is derived.

However, the B&S model hypotheses are rarely appropriate (see e.g. Bates, 2000) and more reliable pricing approaches are often required. As shown by Breeden and Litzenberger (1978), the underlying risk-neutral density is proportional to the second derivative of the option prices taken with respect to the strike prices. Nonetheless, a simple numerical differentiation of the observed option prices is not a feasible solution to the identification problem, since market prices are not available for a continuous set of strikes and are often contaminated by different sources of bias and noise. This has inspired many contributions suggesting different estimation approaches. According to the nature of the hypotheses made about the underlying asset dynamics, we can distinguish three categories of modeling frameworks: parametric, semi-parametric and non-parametric. An extensive discussion of several methods belonging to each class can be found, for example, in Jondeau et al. (2007).

Parametric schemes assume an analytical form for the SPD. For example, Ritchey (1990) models the risk-neutral distribution as a mixture of normal distributions while Bahra (1997) adopts a mixture of log-normal densities.

Semi-parametric methods aim to approximate in a flexible way the departure of the target density from a parametric one, defined according to theoretical argument. An example is the strategy based on Edgeworth expansions (taken around the B&S log-normal SPD) introduced by Jarrow and Rudd (1982). A method based on Hermite polynomials has been proposed by Abken et al. (1996). Jondeau and Rockinger (2001) estimate the unobserved risk-neutral probability law using constrained Gram-Charlier expansions.

Non-parametric approaches do not formulate any hypothesis about the underlying asset dynamics and attempt to estimate the targeted density in a flexible way. Many non-parametric proposals aim an indirect estimation of the SPD based on an optimal approximation/smoothing of the pricing function and a successive approximation of the second derivative of the fitted pay-off. Shimko (1993) suggested to model the observed implied volatility using quadratic polynomials and to derive continuous option pay-offs from the smoothed volatility smiles. Aït-Sahalia and Lo (1998, 2000) and Huynh et al. (2002) proposed kernel smoothing of the observed option prices. Within this curve-fitting framework, Aït-Sahalia and Duarte (2003) recommended an alternative two stages procedure ensuring arbitrage-free estimates: in a first step the data are pre-processed using constrained regression, after which the fitted arbitrage-free prices are smoothed using kernel techniques.

Arbitrage-free estimates can also be obtained using local polynomials. Yatchew and Härdle (2006), for example, designed a smoothing spline approach regularized by a penalty

term forcing estimates compliant to no-arbitrage conditions. Härdle and Hlávka (2009) obtained similar results by using (suitably re-parametrized) constrained non-linear regression but, contrarily to the ones mentioned above, this approach does not ensure smooth estimates of the latent probability density function.

In this work, we introduce a new direct and flexible approach enabling smooth estimates of the state price density implied in option prices (DESPD: direct estimation of the state price density). Our strategy leads to a composite link model (Eilers, 2007), it appears computationally efficient while ensuring arbitrage-free estimates. We treat the observed option prices as expected values of possible pay-offs at maturity weighted by the latent (unknown) density. We model the logarithm of the latter as a smooth function, while matching the expected value of the possible contract's pay-off with the observed prices. The option prices and the related SPD are estimated as functions of optimal regression coefficients obtained by exploiting iterative weighted least squares. In analogy with Eilers and Marx (1996), smoothness is induced by a roughness penalty term allowing for efficient interpolation and extrapolation along unobserved support points (see e.g. Eilers and Marx, 2010).

This paper is organized as follows. In Section 2 we introduce the general settings of our approach and describe the estimation procedure. In Section 3 we discuss the asymptotic properties of the proposed estimators and the consistency of the DESPD estimates with the theoretical no-arbitrage requirements. In Section 4 we evaluate the performance of our proposal through a simulation study while in section 5 we illustrate the results of a real data analyses. Finally, a discussion in Section 6 concludes the paper.

2 DESPD model specification and estimation procedure

Quoting Cox et al. (1979), an option is a contract giving the right (not the obligation) to buy (call type) or sell (put type) a risky asset with price s_T at a predetermined (fixed) strike price k within (American style) or at (European style) a given date (maturity of the contract, T).

Here we consider exclusively European-style (exercise possible only at maturity) option contracts. Intuitively, the current price of the contract should take into account the uncertainty about the pay-off ($s_T - k$) at the expiration date (apart from the cost of money and transaction costs). Under the dynamic equilibrium hypothesis, the fair price of an option contract is equal to the discounted expected value of the possible pay-offs at expiration date

(T) computed under an appropriate density function $f_t(s_T)$: the state price density. Then, at date t , the price of an European call option is given by

$$c_t = \exp(-r_{t,\tau}\tau) \int_0^\infty (s_T - k)^+ f_t(s_T) ds_T, \quad (1)$$

where $r_{t,\tau}$ is the risk-free interest rate (e.g. the Libor rate), $\tau = T - t$ is the time to maturity, s_T (price of the underlying at maturity) is the state variable and $f_t(s_T)$ is the (latent) state price density at time t . As mentioned above, the SPD is not observed directly and must be inferred from quoted option prices.

Define a set of possible underlying asset prices at maturity $\mathbf{u} = \{u_1, \dots, u_m; u_j \in [\min(\mathbf{k}) - \gamma, \max(\mathbf{k}) + \gamma]\}$ (with $\gamma \geq 0$ constant) and let $\varphi_j = f_t(u_j) = \exp(\eta_j)$ where $\boldsymbol{\eta}$ is a $(m \times 1)$ vector of unknowns to be estimated. In analogy with Härdle and Hlávka (2009), assume that $\exp(-r_{t,\tau}\tau) = 1$ (or suppose that the observations are scaled by the known discounting factor). The call option prices can be modeled as

$$c_i = \mu_i + \epsilon_i = \sum_{j=1}^m g_{ij} \varphi_j + \epsilon_i, \quad (2)$$

where $i = 1, \dots, n$, ϵ_i are i.i.d. random variables with zero mean and constant variance σ^2 and $g_{ij} = (u_j - k_i)^+$ (for $i = 1, \dots, n$ and $j = 1, \dots, m$) are the entries of the $(n \times m)$ matrix \mathbf{G} . In principle, the SPD can then be estimated by solving

$$\min_{\boldsymbol{\eta}} \mathcal{S}(\boldsymbol{\eta}) = \|\mathbf{c} - \mathbf{G}\boldsymbol{\varphi}\|^2. \quad (3)$$

Eq. (3) represents a severely ill-conditioned non-linear problem. We regularize it by assuming smoothness of the unknown probability density function and penalize for the differences of the $\boldsymbol{\eta}$ coefficients. This leads to the following penalized non-linear problem (see e.g. Eilers, 2007)

$$\min_{\boldsymbol{\eta}} S(\boldsymbol{\eta}) = \|\mathbf{c} - \mathbf{G}\boldsymbol{\varphi}\|^2 + \lambda \|\mathbf{D}\boldsymbol{\eta}\|^2, \quad (4)$$

where \mathbf{D} is a matrix operator forming third order differences. The parameter λ tunes the smoothness of $\boldsymbol{\varphi}$ and can be selected by means of an optimality criterion (see Section 2.2) or specified by the user.

The non-linear problem in Eq. 4 can be solved via iterative ordinary least squares (for fixed λ). Suppose that an approximation to the mean, $\tilde{\boldsymbol{\mu}}$, is available (in what follows the tilde symbol will always be used to indicate approximation of some unknown). Then, a first

order approximation gives

$$\mu_i \approx \tilde{\mu}_i + \sum_j \frac{\partial \tilde{\mu}_i}{\partial \eta_j} \Delta \eta_j = \tilde{\mu}_i + \sum_j g_{ij} \tilde{\varphi}_j \Delta \eta_j.$$

By combining this result with Eq (4) we can derive the linearized least squares criterion

$$\min_{\boldsymbol{\eta}} \tilde{S}(\boldsymbol{\eta}) = \left\| \mathbf{c} - \tilde{\boldsymbol{\mu}} - \mathbf{G}\tilde{\mathbf{F}}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}) \right\|^2 + \lambda \|\mathbf{D}\boldsymbol{\eta}\|^2, \quad (5)$$

where $\tilde{\mathbf{F}} = \text{diag}(\tilde{\boldsymbol{\varphi}})$. The optimal $\boldsymbol{\eta}$ coefficients can then be estimated through (penalized) iterative weighted least squares, i.e. by solving till convergence the following set of penalized normal equations

$$\left(\tilde{\mathbf{E}}^\top \tilde{\mathbf{E}} + \lambda \mathbf{D}^\top \mathbf{D} \right) \boldsymbol{\eta} = \tilde{\mathbf{E}}^\top \left(\mathbf{c} - \tilde{\boldsymbol{\mu}} + \tilde{\mathbf{E}}\tilde{\boldsymbol{\eta}} \right), \quad (6)$$

with $\tilde{\mathbf{E}} = \mathbf{G}\tilde{\mathbf{F}}$. Convergence of the iterative procedure is usually achieved in a limited number of iterations (less than 30 for a relative tolerance of 10^{-5}).

Eq. (6) can also be modified to deal with heteroscedastic residuals by introducing a suitable set of weights, leading to

$$\left(\tilde{\mathbf{E}}^\top \mathbf{W} \tilde{\mathbf{E}} + \lambda \mathbf{D}^\top \mathbf{D} \right) \boldsymbol{\eta} = \tilde{\mathbf{E}}^\top \mathbf{W} \left(\mathbf{c} - \tilde{\boldsymbol{\mu}} + \tilde{\mathbf{E}}\tilde{\boldsymbol{\eta}} \right).$$

Given that most of the variability is expected for small values of the pay-off, in case of heteroscedasticity, we suggest to set weights equal to the inverse of the ratio between the fitted option prices (as obtained at current iteration of the IWLS procedure) and the observed strikes: $\mathbf{W} = \text{diag}(\tilde{\boldsymbol{\mu}}/\mathbf{k})^{-1}$.

2.1 Including put options

Up to now, we have considered call contracts only. Put options are often traded together with the corresponding calls. The well known put-call parity links the two types of contracts

$$\mathbf{c}(\mathbf{k}, T) = \mathbf{p}(\mathbf{k}, T) + s - \mathbf{k} \exp(-r\tau). \quad (7)$$

By using Eq. (7) one could compute the equilibrium put prices once the value of the related call option has been estimated for different strikes.

Here we propose a different strategy. The model matrix connects the observed prices to the estimated density $\boldsymbol{\varphi}$. By definition, the state price density is unique while \mathbf{G} can be generalized in order to deal with different classes of contracts since its entries depend on the

form of the option's pay-off function. This suggests to use the available put prices as extra observations within the estimation framework presented in the previous section. In other words, we augment the model matrix with row vectors $g_{ij}^* = (k_i - u_j)^+$, the put pay-off, and include the quoted put prices in the vector of observations. The estimation procedure remains the same and the fitted call and put prices are both compliant with no-arbitrage conditions as discussed in Section 3.2.

2.2 Penalty parameter selection

The parameter λ can be selected by means of a suitable optimality criteria. Well-known alternatives are (generalized) cross validation, Akaike's information criterion and the Bayesian information criterion. Figure 1 shows an example based on an AIC minimization. Details about the data used here can be found in Section 5.

We prefer to select the smoothing parameter using an (EM-type) iterative procedure (Schall, 1991) exploiting the link between penalized regression and mixed effects models (see e.g. Lee et al., 2006; Ruppert et al., 2003). Intuitively, λ should ensure an optimal balance between goodness of fit and the degree of smoothing. These two conflicting goals can be summarized by the variance components

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n w_i \left(c_i - \sum_{j=1}^m g_{ij} \hat{\varphi}_j \right)^2}{n - \text{Ed}}, \quad \hat{\sigma}_{\text{PEN}}^2 = \frac{\|\mathbf{D}\hat{\boldsymbol{\eta}}\|^2}{\text{Ed}},$$

with the effective model dimension equal to $\text{Ed} = \text{tr} \left[(\tilde{\mathbf{E}}^\top \mathbf{W} \tilde{\mathbf{E}} + \lambda \mathbf{D}^\top \mathbf{D})^{-1} \tilde{\mathbf{E}}^\top \mathbf{W} \tilde{\mathbf{E}} \right]$ (see e.g. Hastie and Tibshirani, 1990). This suggests to iteratively update the smoothing parameter defined as $\lambda = \hat{\sigma}^2 / \hat{\sigma}_{\text{PEN}}^2$. Convergence is achieved in a small number of iterations (usually less than 15). Under the normality assumption, a strict relationship connects the AIC optimization and the EM-like approach, since the λ selected by the latter method can be shown to be equal to the expected value of the one suggested by the AIC criterion (see Krivobokova and Kauermann, 2007).

3 Asymptotic and arbitrage-free properties

In this section we discuss the asymptotic and arbitrage properties of the DESPD estimators described above. The proofs of the following asymptotic results can be found in Appendix A.

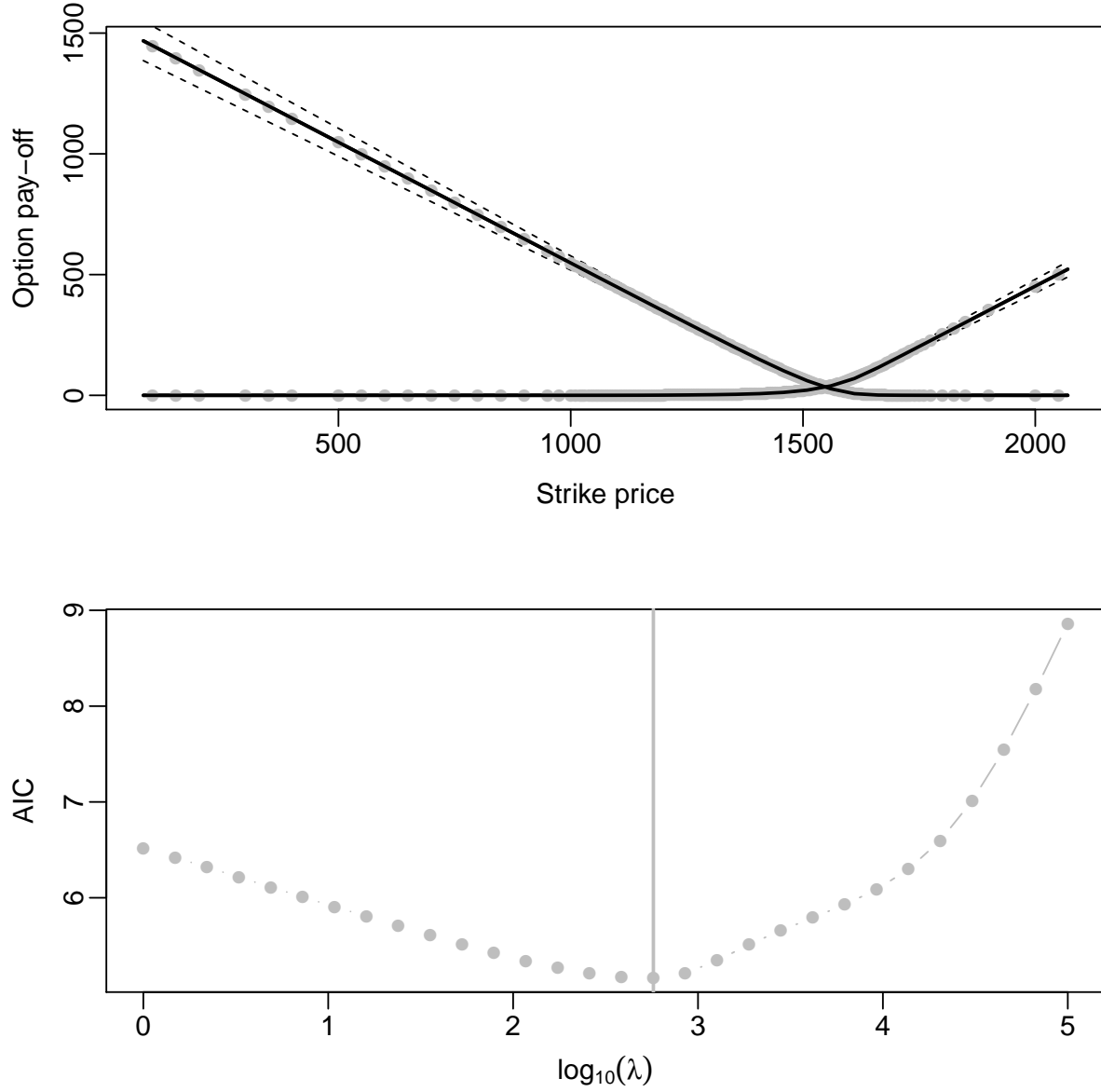


Figure 1. Upper panel: observed put and call option prices (gray dots), smooth pay-offs (solid black lines) and pointwise confidence bounds (gray dashed lines). Lower panel: AIC values for different values of $\log_{10}(\lambda)$. The vertical line indicates the minimum of the selection criterion.

In order to study the asymptotic behavior of the proposed estimators, it is convenient to reformulate Eq. (4) as

$$\min_{\boldsymbol{\eta}} S_n(\boldsymbol{\eta}) = n^{-1} \sum_{i=1}^n (c_i - \mathbf{g}_i^\top \boldsymbol{\varphi})^2 + \lambda_n \|\mathbf{D}\boldsymbol{\eta}\|^2, \quad (8)$$

where \mathbf{g}_i is a $(m \times 1)$ vector with entries equal to i th row of the model matrix and the index n in λ_n emphasizes the influence of the sample size on the smoothing parameter. The one in Eq. (8) is a penalized regression linear in the unknown state price density and nonlinear in the $\boldsymbol{\eta}$ parameters and represents a special case of the constrained coefficients model discussed in Malinvaud (1970, par. 3).

3.1 Asymptotics for $\lambda_n \rightarrow 0$

Suppose that $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$ and that the following assumptions are satisfied

Ass 1 The parameter space \mathcal{E} is compact and includes the true parameters $\boldsymbol{\eta}^0 \in \mathcal{E}$.

Ass 2 For n sufficiently large, the second-order moment matrix $\mathcal{G}_{gg} = n^{-1} \sum_{i=1}^n \mathbf{g}_i \mathbf{g}_i^\top$ converges to a nonsingular matrix $\bar{\mathcal{G}}$ for $\boldsymbol{\eta}$ in a neighborhood of $\boldsymbol{\eta}^0$.

Ass 3 It is verified that

$$\mathcal{K}(\boldsymbol{\eta}^0) = \lim_n n^{-1} \sum_{i=1}^n \frac{\partial \mu_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \bigg|_{\boldsymbol{\eta}^0} \frac{\partial \mu_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^\top} \bigg|_{\boldsymbol{\eta}^0}$$

exists and is nonsingular and that

$$n^{-1} \sum_{i=1}^n \frac{\partial \mu_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\partial \mu_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^\top} \text{ and } n^{-1} \sum_{i=1}^n \frac{\partial^2 \mu_i(\boldsymbol{\eta})}{\partial \eta_j \partial \eta_l}$$

converge uniformly in $\boldsymbol{\eta}$ in a neighborhood of $\boldsymbol{\eta}^0$ for all $j, l \in \dim(\boldsymbol{\eta})$.

PROPOSITION 1. *Under Ass 1-2 and for $\lambda_n = o(1)$ a sequence of estimators $\boldsymbol{\eta}$ satisfying Eq. (8) exists and is a consistent estimator of $\boldsymbol{\eta}^0$.*

PROPOSITION 2. *Under Ass 1-3 and for $\lambda = o(n^{-1/2})$, a sequence of penalized non linear least squares estimators $\boldsymbol{\eta}$ exists, is consistent and is asymptotically normally distributed.*

The proofs of theorems 1 and 2 can be found in Appendix A. When convergence of the iterative least squares estimation has been reached for a given value of λ , the variance-covariance matrix of $\hat{\boldsymbol{\eta}}$ is equal to

$$\text{Var}(\hat{\boldsymbol{\eta}}) = \hat{\sigma}^2 \left(\tilde{\mathbf{E}}^\top \tilde{\mathbf{E}} + \lambda \mathbf{D}^\top \mathbf{D} \right)^{-1}. \quad (9)$$

The variance of the estimated SPD can be obtained from Eq. (9) using the delta method.

3.2 No-arbitrage properties of the estimates

The estimated option prices must be consistent with no-arbitrage conditions. Here we discuss the arbitrage properties of the DESPD estimates. Consider for the moment call prices only. According to Harrison and Pliska (1981) and Härdle and Hlávka (2009), the following properties must hold for arbitrage-free estimates

1. The estimated density is proper: $\hat{\varphi}_j \geq 0$, $\forall j = 1, \dots, m$ and $\sum_{j=1}^m \hat{\varphi}_j = 1$,
2. The estimated prices are non-negative: $\hat{c}_i \geq 0$, $\forall i = 1, \dots, n$,
3. The estimated pricing function is monotonically decreasing: $\frac{\partial \hat{c}_i}{\partial k_i} \leq 0$, $\forall i = 1, \dots, n$,
4. The estimated pricing function is convex: $\frac{\partial^2 \hat{c}_i}{\partial k_i^2} \geq 0$, $\forall i = 1, \dots, n$.

The positivity requirement in condition 1 is satisfied since, by definition, $\hat{\varphi}_j = \exp(\hat{\eta}_j) \geq 0$, $\forall \hat{\eta}_j \in \mathbb{R}$. The sum-to-one property of $\hat{\varphi}$ can be introduced as an explicit constraint. In analogy with what described in Currie (2013), the constrained penalized optimization problem can be written as

$$\min_{\boldsymbol{\eta}} S_c(\boldsymbol{\eta}) = \|\mathbf{c} - \mathbf{G}\boldsymbol{\varphi}\|^2 + \lambda \|\mathbf{D}\boldsymbol{\eta}\|^2 + \omega (\mathbf{1}^\top \boldsymbol{\varphi} - 1),$$

where $\mathbf{1}$ is a m -dimensional vector of ones and ω is a Lagrange multiplier. The nonlinear condition can be approximated for given values of the unknowns ($\tilde{\boldsymbol{\eta}}$) by using first order Taylor expansion as $\mathbf{1}^\top \boldsymbol{\varphi} \approx \mathbf{1}^\top \tilde{\boldsymbol{\varphi}} + \tilde{\boldsymbol{\varphi}}^\top (\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})$. The optimal $\boldsymbol{\eta}$ vector can then be estimated by solving iteratively the linear problem

$$\min_{\boldsymbol{\eta}} \tilde{S}_c(\boldsymbol{\eta}) = \left\| \mathbf{c} - \tilde{\boldsymbol{\mu}} - \tilde{\mathbf{E}}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}) \right\|^2 + \lambda \|\mathbf{D}\boldsymbol{\eta}\|^2 + \omega (\mathbf{1}^\top \tilde{\boldsymbol{\varphi}} + \tilde{\boldsymbol{\varphi}}^\top (\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}) - 1),$$

where $\tilde{\boldsymbol{\mu}} = \mathbf{G}\tilde{\boldsymbol{\varphi}}$, $\tilde{\mathbf{E}} = \mathbf{G}\tilde{\mathbf{F}}$ and $\tilde{\mathbf{F}} = \text{diag}(\tilde{\boldsymbol{\varphi}})$. This leads to the system of constrained penalized normal equations (which can be generalized to include put prices and error heteroscedasticity, as discussed above)

$$\begin{bmatrix} \tilde{\mathbf{E}}^\top \tilde{\mathbf{E}} + \lambda \mathbf{D}^\top \mathbf{D} & \tilde{\boldsymbol{\varphi}} \\ \tilde{\boldsymbol{\varphi}}^\top & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \omega \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{E}}^\top (\mathbf{c} - \tilde{\boldsymbol{\mu}} + \tilde{\mathbf{E}}\tilde{\boldsymbol{\eta}}) \\ 1 - \mathbf{1}^\top \tilde{\boldsymbol{\varphi}} + \tilde{\boldsymbol{\varphi}}^\top \tilde{\boldsymbol{\eta}} \end{bmatrix}.$$

Condition 2 is clearly satisfied by the definition of the mean function in Eq. (2). In order to prove conditions 3 and 4 it is convenient to express the i th estimated call price as

$$\hat{c}_i = \sum_{j=1}^m (u_j - k_i) \mathbf{H}(u_j - k_i) \hat{\varphi}_j,$$

where $\mathbf{H}(u_j - k_i)$ is an Heaviside step function with value one if $u_j \geq k_i$ and zero otherwise. Define now the delta function $\delta(u_j - k_i)$ equal to zero for $u_j \neq k_i$. Computing the first and second derivatives (w.r.t. k_i) of the estimated option prices we obtain

$$\begin{aligned} \frac{\partial \hat{c}_i}{\partial k_i} &= - \sum_{j=1}^m \mathbf{H}(u_j - k_i) \hat{\varphi}_j - \sum_{j=1}^m (u_j - k_i) \delta(u_j - k_i) \hat{\varphi}_j = - \sum_{j=1}^m \mathbf{H}(u_j - k_i) \hat{\varphi}_j \\ &\approx - \int_k^\infty \hat{\varphi}(x) dx \leq 0, \\ \frac{\partial^2 \hat{c}_i}{\partial k_i^2} &= \sum_{j=1}^m \delta(u_j - k_i) \hat{\varphi}(u_j) \approx \hat{\varphi}(k_i) \geq 0. \end{aligned}$$

Finally, conditions 1 and 2 clearly hold also for the put prices estimated by using a similar least squares procedure. Conditions 3 (with a positive sign now) and 4 can be demonstrated in analogy with what shown for call options since

$$\hat{p}_i = \sum_{j=1}^m (k_i - u_j) \mathbf{H}(k_i - u_j) \hat{\varphi}_j.$$

4 Simulation analysis

In this section we test our approach through a simulation study. We simulate 500 plain vanilla European options under the B&S model, which allows an objective evaluation of the performance achieved by the DEPSD estimates, since the risk-neutral p.d.f. is known a priori. Indeed, within this framework, the targeted distribution is log-normal with mean $\mu^* = (r - 0.5\varsigma^2)\tau$ and standard deviation $\sigma^* = \varsigma\sqrt{\tau}$ where ς is the implied volatility parameter and $\tau = T - t$ is the residual life of the contract.

We have generated option contracts taking $n = \{10, 20, 30\}$ equally spaced strike prices between \$40 and \$140 with a spot price $s = \$75$, maturity in $\tau = 30$ days, constant implied volatility $\sigma = 0.2$, risk-free rate $r = 3\%$ and a divided yield equal to zero. In order to mimic miss-pricing effects, in analogy with what suggested by Aït-Sahalia and Duarte (2003), we assume a bid-ask spread of 3% with a floor of 30 cents and a cap of 2.5 dollars and add to the simulated prices a noise term drawn from a uniform distribution defined between 0% and

50% of the bid-ask spread. Finally, we simulate different degrees of moneyness by assuming most of the liquidity near the money. This is achieved by multiplying the error term by the liquidity factor $1 + 5|M_{t,\tau} - 1|$ with $M_{t,\tau} = k/(s_t \exp(\tau r))$.

We evaluate the DEPSD estimates by looking at: 1) the goodness of fit of the fitted option prices and the appropriateness of the estimated density, 2) the forecasting ability of the inferred SPDs. In particular, we assess point 1) by computing over the simulation runs

- the integrated squared error (ISE) between the theoretical (f^*) and estimated ($\hat{\varphi}$) SPDs

$$\text{ISE} = \int_0^\infty (f^*(x) - \hat{\varphi}(x))^2 dx,$$

- the (absolute) relative bias of the estimated SPD means and standard deviations

$$\text{Bias}_\mu = \left| \frac{\mu^* - \hat{\mu}}{\mu^*} \right|, \quad \text{Bias}_\sigma = \left| \frac{\sigma^* - \hat{\sigma}}{\sigma^*} \right|,$$

- the root mean squared errors between the theoretical prices (\mathbf{y}^*) and the estimated ones ($\hat{\mathbf{y}}$)

$$\text{RMSE} = \sqrt{\sum_{i=1}^{2n} \frac{(y_i^* - \hat{y}_i)^2}{2n}}.$$

Figure 2 summarizes our results. The integrated squared errors between the estimated and the theoretical SPDs are really small for all the sample (strike) sizes. The appropriateness of our estimates can also be evaluated by looking at the size of the relative bias for the estimated means and standard deviations (second and third panel of Figure 2). In addition, the root mean squared errors between the theoretical and estimated prices are (on average) quite moderate highlighting the good pricing properties of our estimates. Finally, the observed performance indicators show a moderate variability which, as expected, tends to decrease for larger sample sizes.

The forecast ability of DESPD can be evaluated using probability integral transforms (PIT, Rosenblatt, 1952). The idea behind the PIT-based assessment can be summarized as follows. Suppose that at time t the density function $\hat{\varphi}_{t+\Delta t}$, i.e. the SPD at Δt days ahead, is available. Suppose also that we know the spot price $s_{t+\Delta t}$. Then we can define the quantity

$$z_{t+\Delta t} = \int_{-\infty}^{s_{t+\Delta t}} \hat{\varphi}_{t+\Delta t}(x) dx = \hat{\mathcal{F}}_s(s_{t+\Delta t}),$$

where $\mathcal{F}(\cdot)$ indicates a cumulative distribution function so that $z_{t+\Delta t}$ is equal to the probability of $s_{t+\Delta t}$ under the density $\hat{\varphi}_{t+\Delta t}$. Assume that the estimated SPD is the true density

of $s_{t+\Delta t}$, then

$$\begin{aligned}
\mathcal{F}_z(z_{t+\Delta t}) &= \Pr(z \leq z_{t+\Delta t}) \\
&= \Pr\left(\hat{\mathcal{F}}_s(s_{t+\Delta t}) \leq z_{t+\Delta t}\right) \\
&= \Pr\left(s_{t+\Delta t} \leq \hat{\mathcal{F}}_s^{-1}(z_{t+\Delta t})\right) \\
&= \hat{\mathcal{F}}_s\left(\hat{\mathcal{F}}_s^{-1}(z_{t+\Delta t})\right) = z_{t+\Delta t}.
\end{aligned}$$

Hence $z_{t+\Delta t}$ follows a $U(0, 1)$ distribution if the estimated SPD is the true density of $s_{t+\Delta t}$ and the forecast ability of the DESPD approach can be evaluated by comparing the distribution of the $z_{t+\Delta t}$ under the ex-ante SPD with the theoretical uniform one. Following Berkowitz (2001), we prefer here to normalize the PIT measure as $x_{t+\Delta t} = \Phi^{-1}(z_{t+\Delta t})$, with $\Phi(\cdot)$ indicating the (standardized) Gaussian distribution function.

Figure 3 shows the quantile-quantile plots of the normalized PITs obtained with $\Delta t = 30$ and 60 days for the three sample sizes used in the simulation study. The future spot prices have been simulated using a geometric Brownian motion dynamics for s_t (in consistency with the settings summarized above). From our results we can conclude that the estimated SPDs predict well the body of the true density for all simulation settings even if the accuracy of the forecast seems slightly lower for smaller sample sizes in correspondence of the most extreme quantiles.

5 Real data examples

In this section we test DESPD by dealing with three historical index option contracts: two written on the S&P 500 index and one on the CBOE Volatility Index. These data are part of the R package RND (Hamidieh, 2014).

Index options are often used to estimate the state price density. European-style contract written on the S&P 500 index are, for example, the most actively traded financial derivatives and, given the nature of the underlying asset, they represent the contract for which most likely the assumptions of the B&S model could be verified (as discussed also by Aït-Sahalia and Lo, 1998). In contrast, VIX options are non-equity instruments having implied volatilities as underlying. The SPD implied by this kind of contract provides important information about volatility market expectations and can be used, for example, to estimate time-varying risk premia (see e.g. Bollerslev and Todorov, 2011).

In our analyses, we compare the DESPD estimates with those achieved by two competitors: the model proposed by Bahra (1997) based on a mixture of B&S log-normal densities

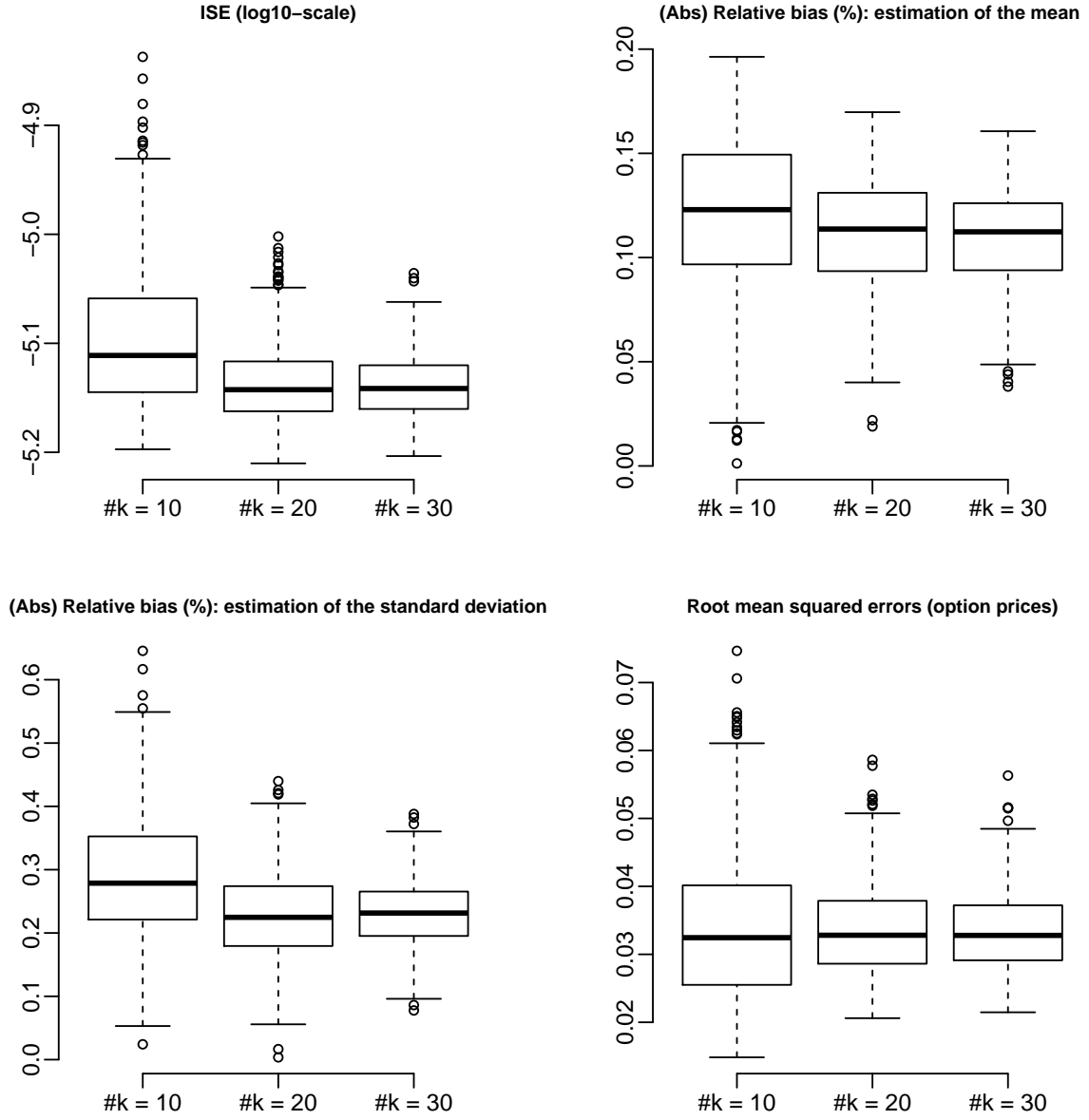


Figure 2. Boxplots of the ISE, relative bias of the estimated mean of the SPD, relative bias of the estimated standard deviation of the SPD and RMSE of the fitted option prices. These results have been obtained for strike price vectors with increasing dimension ($\#k$).

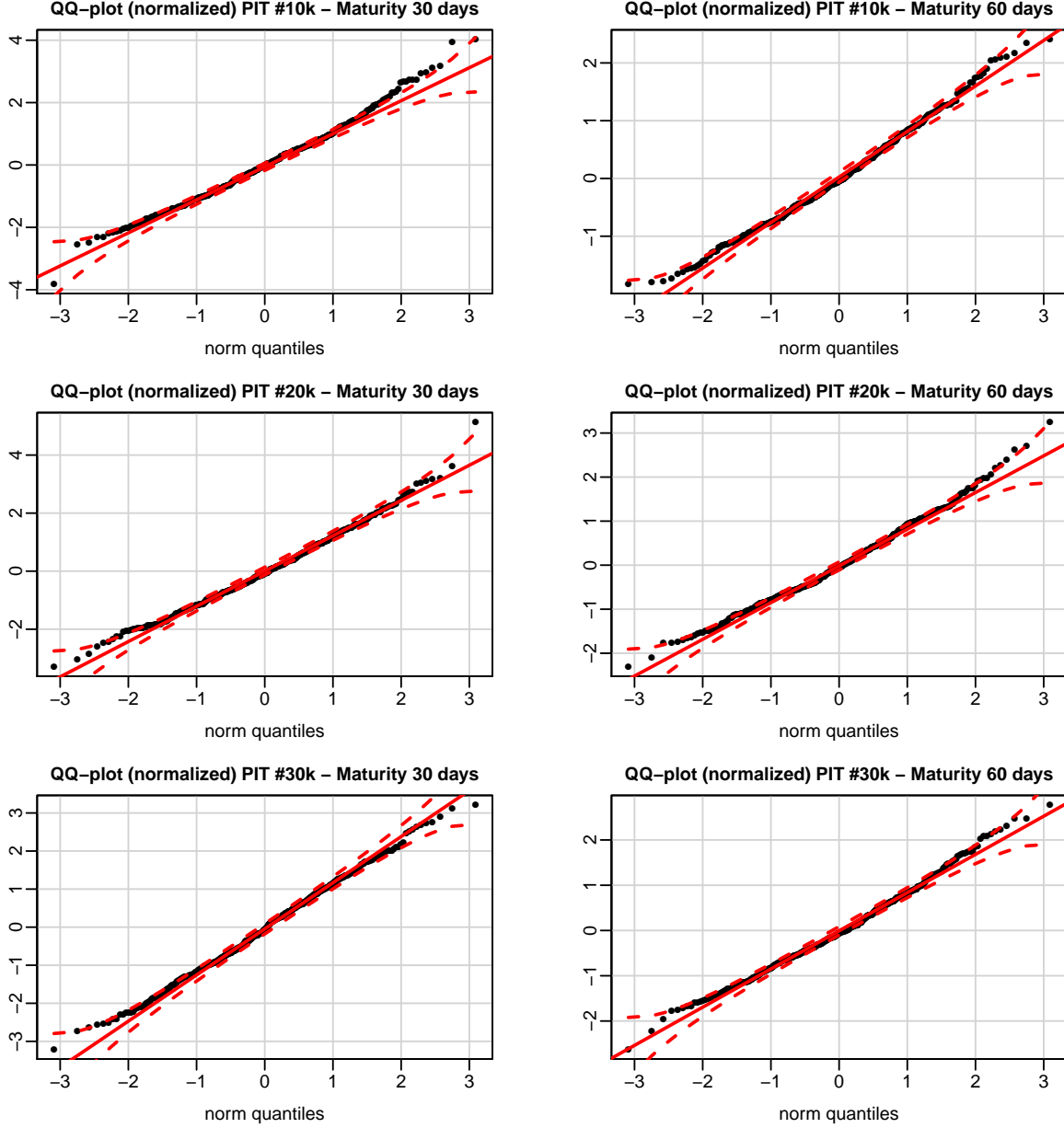


Figure 3. Quantile-quantile plots of normalized probability integral transforms for different time horizons and strike price vectors with increasing dimension ($\#k$). The empirical quantiles (black dots) are compared with the theoretical normal ones (solid red line). The confidence envelopes (dashed red lines) are based on the standard errors of the order statistics of an independent random sample drawn from a standard normal distribution.

and the semi-parametric Edgeworth expansion introduced by Jarrow and Rudd (1982). A mixture of log-normals is one of the most flexible parametric alternatives for the estimation of the SPD. The unobserved density is described by a five parameter log-normal mixture model. The option prices (the expected values of the pay-off at maturity under the hypothesized density) are obtained as linear combinations of B&S prices computed for different shape and scale parameters. The unknowns are estimated by minimizing the squared deviations between the observed and the estimated prices.

The approach of Jarrow and Rudd (1982) is slightly different but still takes the B&S framework as reference. The idea is to model the deviations from log-normality using Edgeworth expansions. Once the SPD has been obtained for a given set of unknowns, the option prices are computed by integration as in Eq. (1). The unknown parameters are once again estimated by minimizing the squared deviations between market (observed) and estimated prices.

As first examples, we analyse two European options written on the S&P 500 index with maturities in 62 and 53 days observed on April 19, 2013 and on June 24, 2013 respectively. The DESPD estimates are presented in Figures 4 and 6, while the results obtained using the two alternative methods can be found in Figures 5 and 7. For the option contract expiring in 63 days, the DESPD and the mixture of log-normals estimates look quite similar, although the latter shows a small “hill” in the proximity of strike 1400. The root mean squared errors (RMSE) were equal to 0.411, 0.567 and 4.389 for the DESPD, mixture-model and Edgeworth expansion methods respectively. Notice that the RMSE for the DESPD estimates has been evaluated using a leave-one-out cross-validation (LOOCV) in order to take into account for possible data overfitting.

The same conclusions can be drawn for the contract with maturity in 53 days. In this case the pricing root mean squared errors were equal to 0.271 (LOOCV), 0.621 and 3.186 respectively. In this case, the density estimated via Edgeworth expansion shows negative values in the right tail, a known drawback of this class of methods.

As a third example, we analyse the historical VIX option prices observed on June 25, 2013 for a contract with maturity in 57 days. The results are presented in Figure 8 and 9. In this case Edgeworth expansion did not give reasonable estimates (not shown). The mixture of log-normals and the DESPD results look quite different. However, the latter provides more accurate price estimates ensuring a (LOOCV-based) root mean squared error equal to 0.025 (the one for the mixture of log-normals was found equal to 0.056).

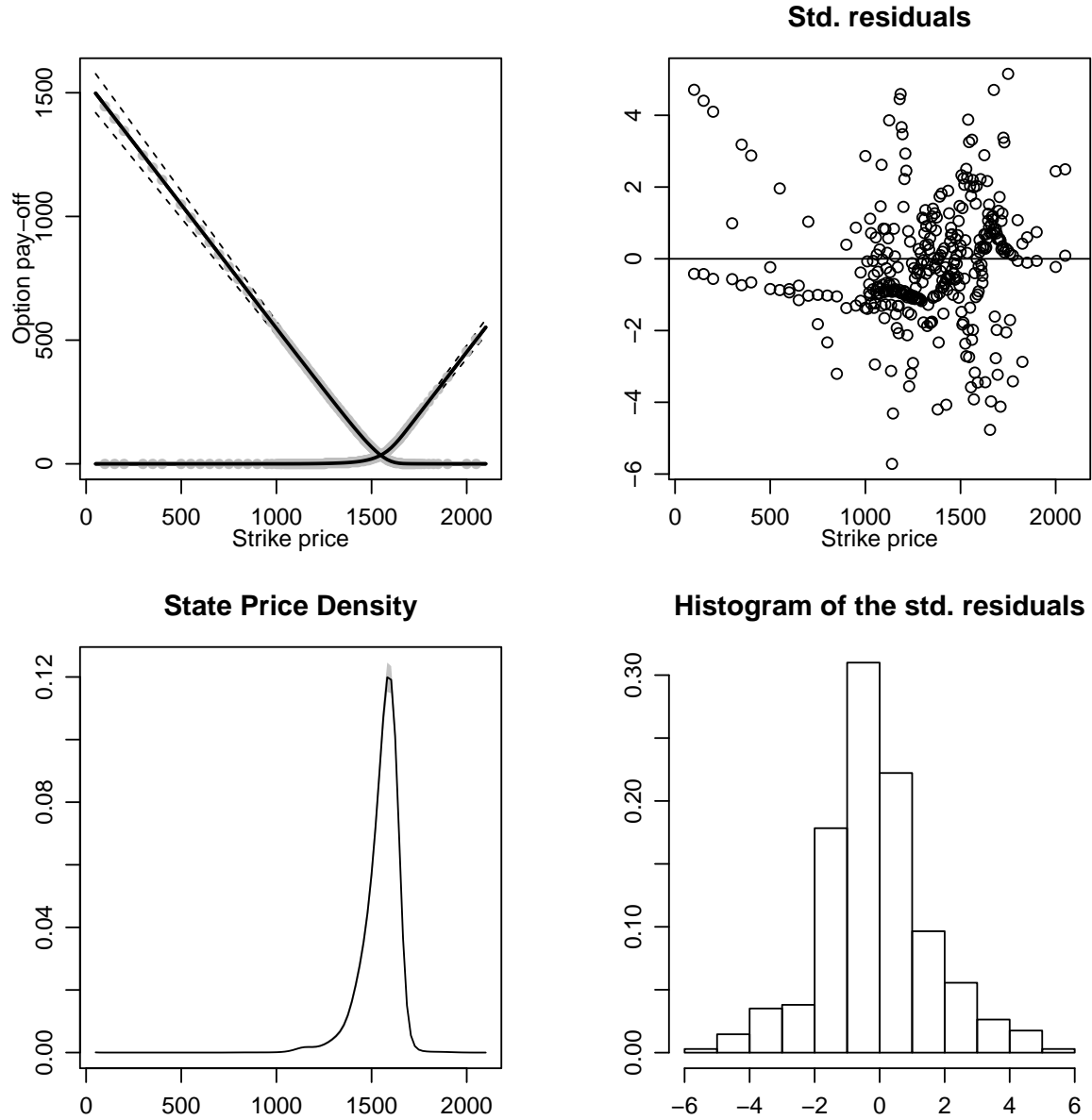


Figure 4. DESPD estimates for the S&P option prices with maturity 62 days. The option prices (gray dots) have been observed on April 19, 2013. The optimal smoothing parameter was found equal to 474.62.

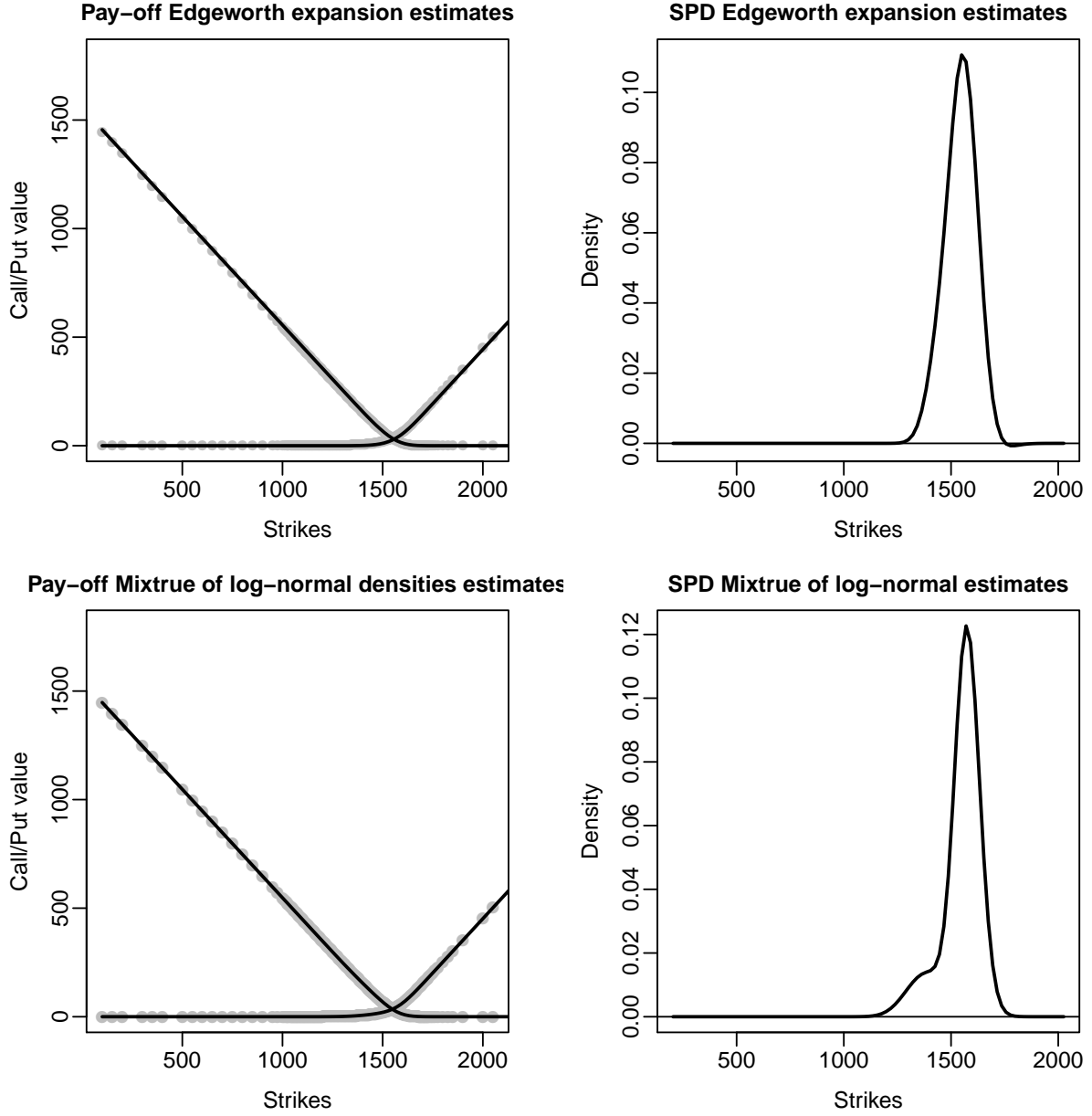


Figure 5. Estimates for the S&P option prices (gray dots) with maturity 62 days obtained by using mixture of log-normal densities and the Edgeworth expansion method.

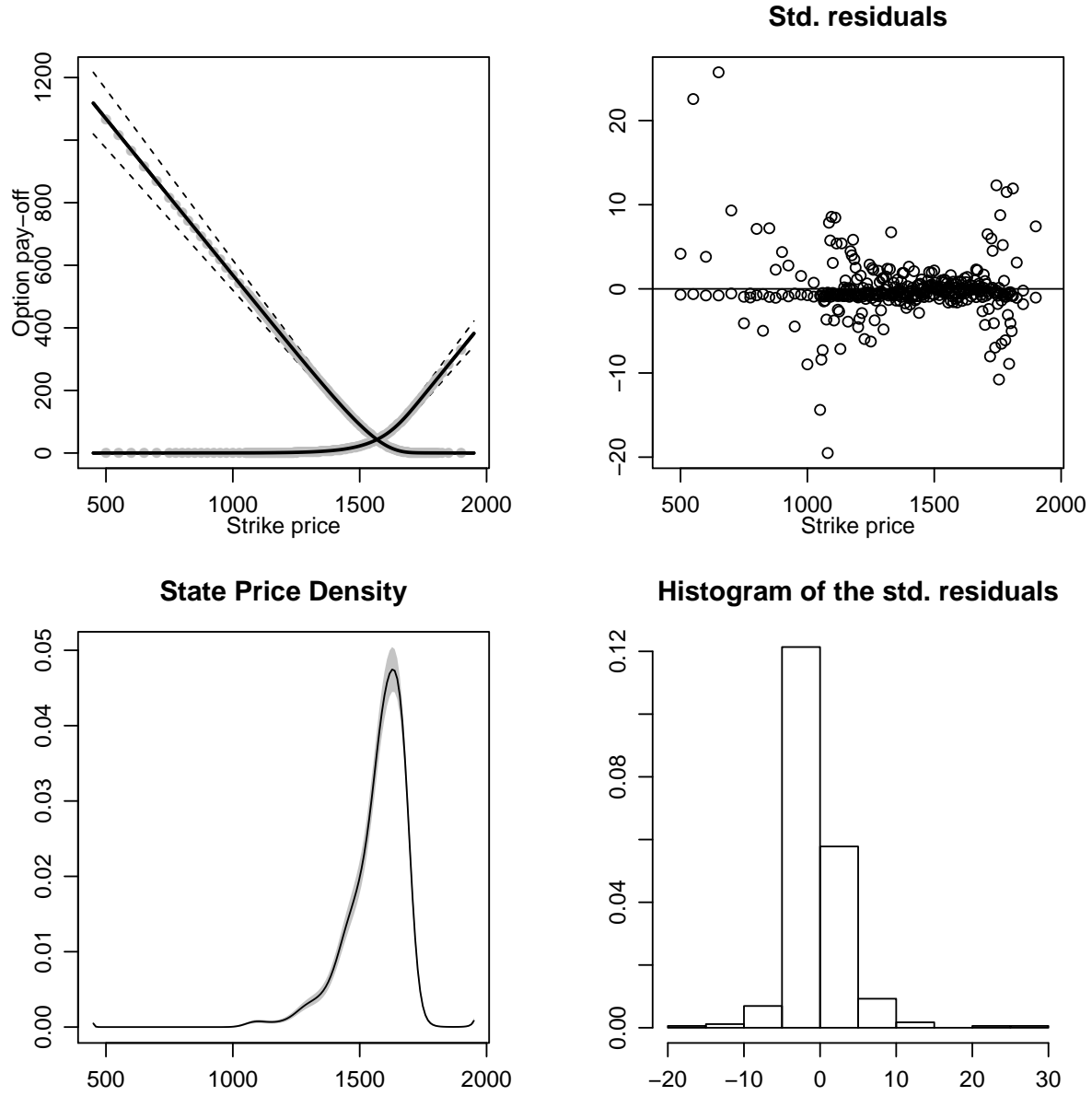


Figure 6. DESPD estimates for the S&P option prices (gray dots) with maturity 53 days observed on June 24, 2013. The optimal smoothing parameter was found equal to 504.24.

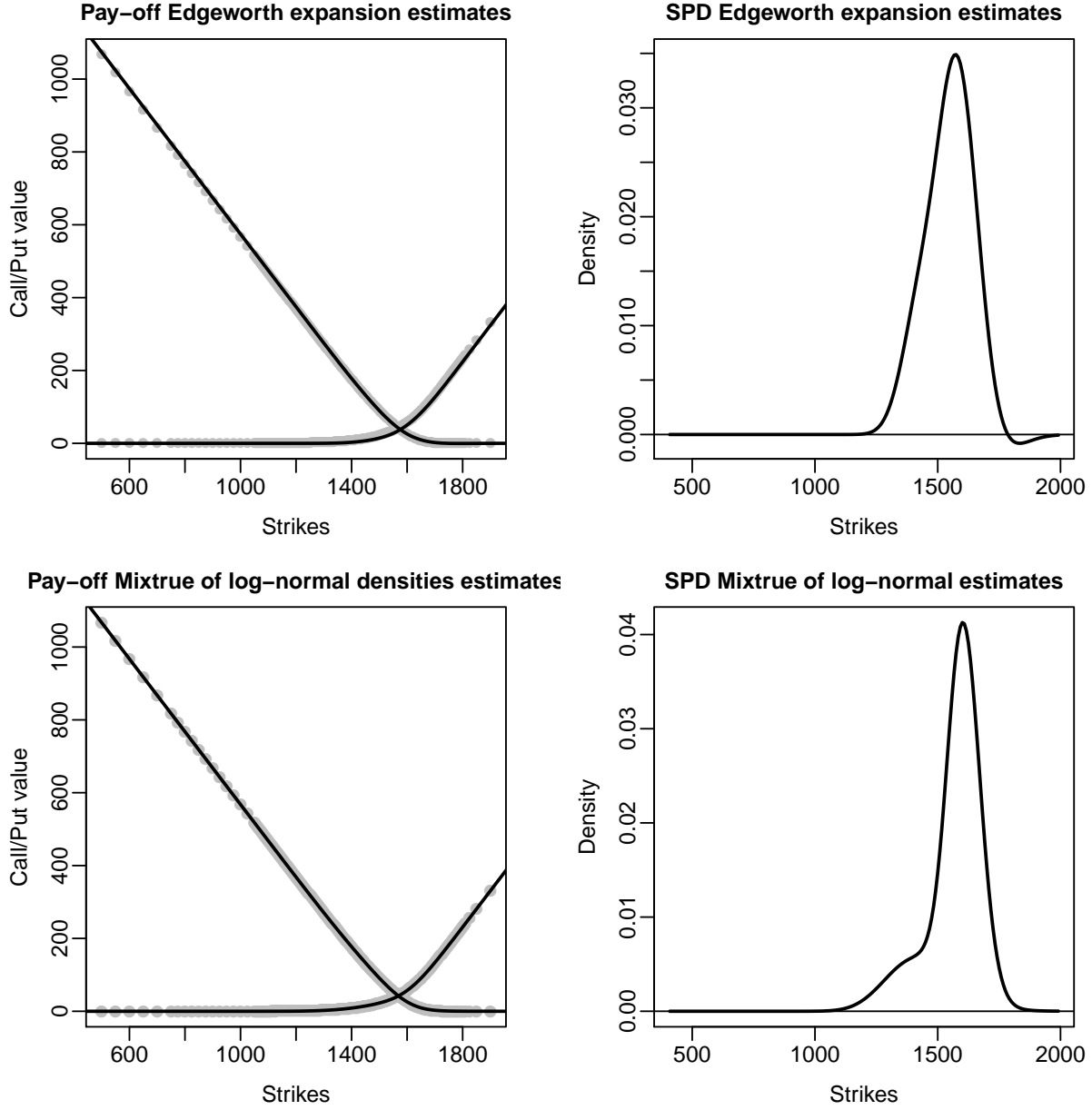


Figure 7. Estimates for the S&P option prices (gray dots) with maturity 53 days obtained by using mixture of log-normal densities and the Edgeworth expansion method.

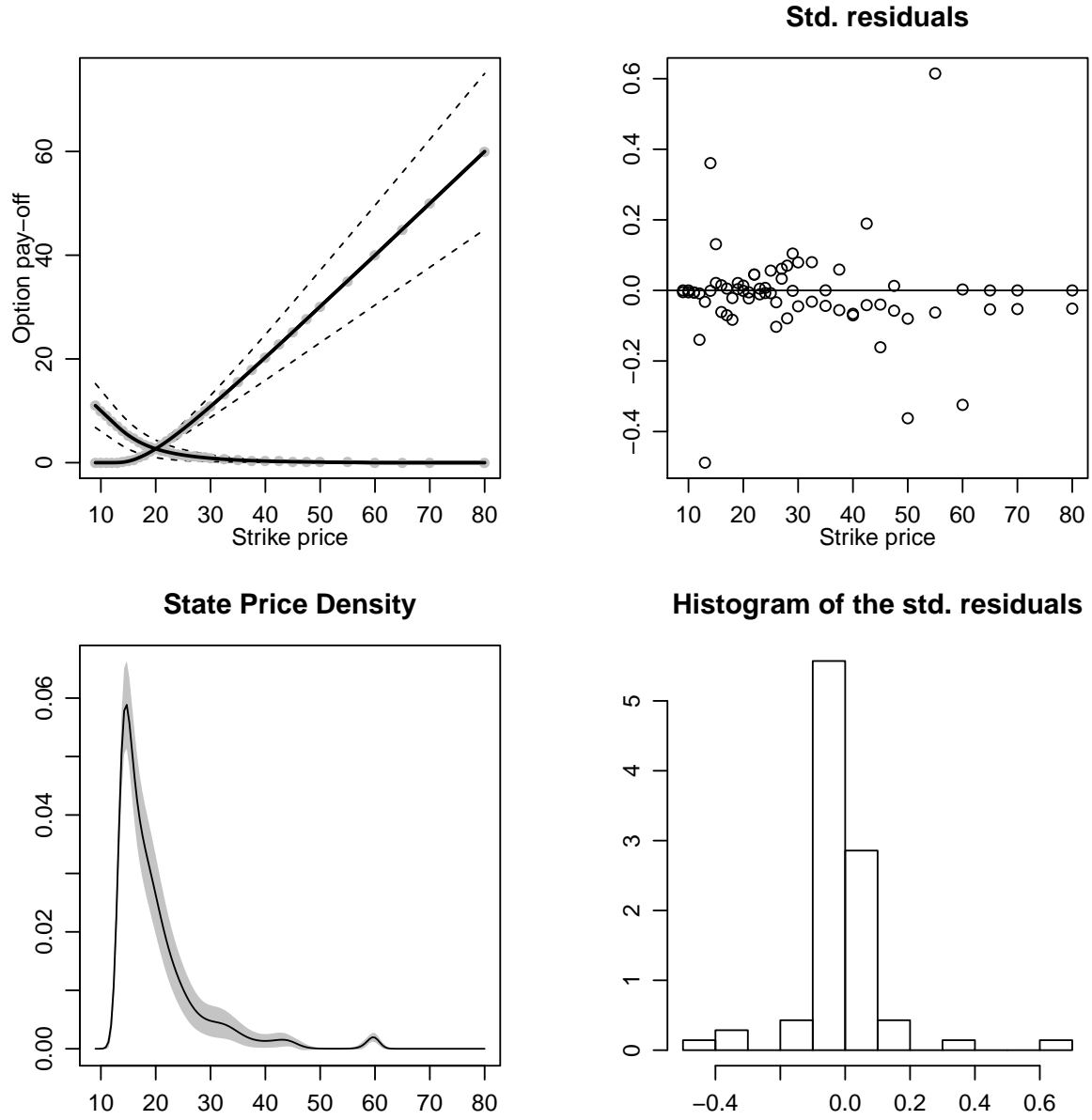


Figure 8. DESPD estimates for the VIX option prices (gray dots with maturity 57 days. The optimal λ parameter was found equal to 11.97.

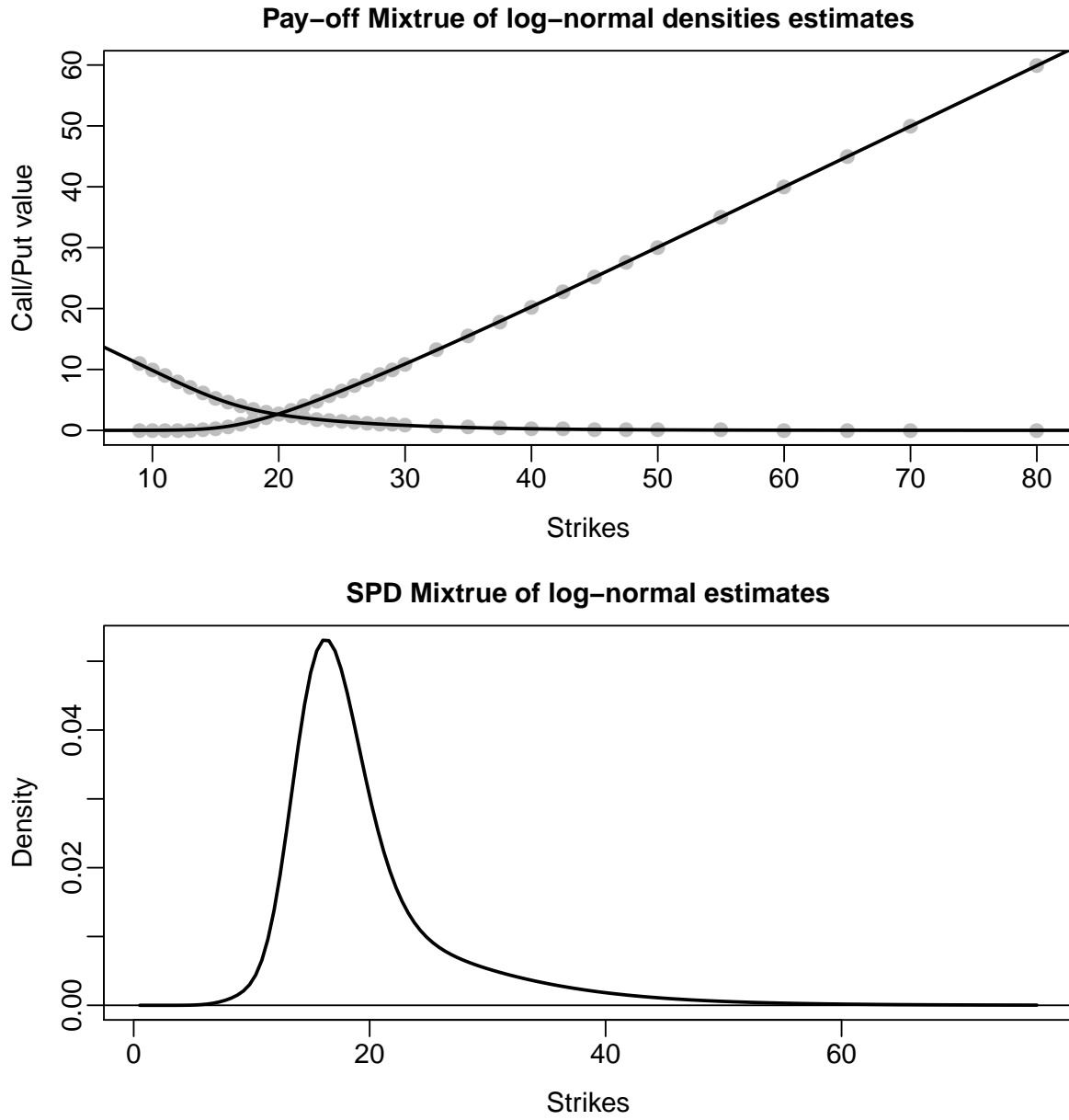


Figure 9. Estimates for the for the VIX option prices (gray dots) with maturity 57 days obtained by using mixture of log-normal densities.

6 Discussion

We have introduced a new semi-parametric method for the direct estimation of the state price density (SPD) implied in option prices (DESPD). We define the observed option prices as expected values of possible payoffs at maturity under the unknown equilibrium density. In particular we model the logarithm of the SPD as a smooth function and match the expected values of the possible pay-offs at maturity with the observed prices. This problem is ill-conditioned. Therefore, we regularize it by using a roughness penalty ensuring smooth estimates of the unknown SPD and efficient extrapolation to unobserved pay-offs. Our procedure can be seen as a special case of the penalized composite link model (Eilers, 2007). The asymptotic properties of the proposed estimators have been presented in Section 3.1.

Our approach is particularly flexible and accurate, does not rely on any parametric assumption about the dynamics of the option's underlying asset and ensures arbitrage-free estimates (see Section 3.2). The SPD and the expected option prices are estimated by iterative weighted least squares (IWLS).

We have tested the DESPD performances through simulations and real data analyses. Our Monte Carlo experiments compared the DESPD estimates with the theoretical option values obtained within the Black and Scholes (1973) framework. This offered a certain benchmark for the evaluation of the merits of our modeling strategy. The results presented in Section 4 highlight the quality of our estimates in terms of pricing accuracy, density estimation and forecast ability.

In Section 5 we tested our methodology on historical index option contracts. We compare the estimated prices and SPDs with those obtained by two alternative methods: the optimal Edgeworth expansion around the B&S log-normal density proposed by Jarrow and Rudd (1982), and the framework presented by Bahra (1997) based on an optimal mixture of log-normal densities. From our results it appears that the DESPD model outperforms the competitors.

Our future research will focus on the generalization the DESPD framework. Using tensor product P-splines, two-dimensional SPDs can be estimated. The second dimension could be the time to maturity or the hours of the trading day.

References

- Abken, P. A., Madan, D. B., and Ramamurtie, B. S. (1996). Estimation of risk-neutral and statistical densities by hermite polynomial approximation: with an application to eurodollar futures options. FRB Atlanta Working Paper 96-5, Federal Reserve Bank of Atlanta.
- Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1-2):9–47.
- Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53(2):499–547.
- Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1-2):9–51.
- Amemiya, T. (1983). Non-linear regression models. *Handbook of Econometrics*, pages 333 – 389. Elsevier.
- Bahra, B. (1997). Implied risk-neutral probability density functions from option prices: theory and application. Bank of England working papers 66, Bank of England.
- Bates, D. S. (2000). Post-'87 crash fears in the s&p 500 futures option market. *Journal of Econometrics*, 94(1-2):181–238.
- Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637.
- Bollerslev, T. and Todorov, V. (2011). Tails, Fears, and Risk Premia. *Journal of Finance*, 66(6):2165–2211.
- Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–51.
- Cox, J., Ross, S., and Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263.
- Cox, J. C. and Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3(12):145 – 166.

- Currie, I. D. (2013). Smoothing constrained generalized linear models with an application to the lee-carter model. *Statistical Modelling*, 13(1):69–93.
- Eilers, P. H. (2007). Ill-posed problems with counts, the composite link model and penalized likelihood. *Statistical Modelling*, 7(3):239–254.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121.
- Eilers, P. H. C. and Marx, B. D. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6):637–653.
- Hamidieh, K. (2014). *RND: Risk Neutral Density Extraction Package*. R package version 1.1.
- Härdle, W. and Hlávka, Z. (2009). Dynamics of state price densities. *Journal of Econometrics*, 150(1):1–15.
- Harrison, J. and Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3):381 – 408.
- Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. London: Chapman & Hall.
- Huynh, K. P., Kervella, P., and Zheng, J. T. (2002). Estimating state-price densities with nonparametric regression. SFB 373 Discussion Papers 2002,40, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.
- Jarrow, R. and Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, 10(3):347–369.
- Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40(2):633–643.
- Jondeau, E., Poon, S.-H., and Rockinger, M. (2007). *Financial modeling under non-gaussian distributions*. Springer Finance. Springer.
- Jondeau, E. and Rockinger, M. (2001). Gramcharlier densities. *Journal of Economic Dynamics and Control*, 25(10):1457 – 1483.

- Krivobokova, T. and Kauermann, G. (2007). A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*, 102:1328–1337.
- Lee, Y., Nelder, J., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.
- Malinvaud, E. (1970). The consistency of nonlinear regressions. *Ann. Math. Statist.*, 41(3):956–969.
- Ritchey, R. J. (1990). Call option valuation for discrete normal mixtures. *Journal of Financial Research*, 13(4):285–296.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.
- Ruppert, D., Wand, P., and Carroll, R. (2003). *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78(4):719–727.
- Shimko, D. (1993). Bounds of probability. *RISK*, 6:33–47.
- Yatchew, A. and Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, 133(2):579–599.
- Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97:1042–1054.

Appendix A: Proofs of the results of Section 3.1

Proof of Proposition 1

The existence of the estimator $\boldsymbol{\eta}$ follows from the results presented in Jennrich (1969) given that the mean function $\mu_i(\boldsymbol{\eta}) = \boldsymbol{\varphi}^\top \mathbf{g}_i = \exp(\boldsymbol{\eta})^\top \mathbf{g}_i$ is continuous for all $\boldsymbol{\eta} \in \mathcal{E}$ and that \mathcal{E} is compact by assumption 1. Here we aim to prove the consistency of this estimator.

The penalized least squares estimator $\hat{\boldsymbol{\eta}}$ is the one minimizing Eq. (8). The term $\lambda_n \|\mathbf{D}\boldsymbol{\eta}\|^2$ tends to zero since, by assumption, $\lambda_n = o(1)$, $\boldsymbol{\eta} \in \mathcal{E}$ and \mathcal{E} is compact. Define now the quantities

$$\begin{aligned} R_n(\boldsymbol{\eta}) &= n \left(\exp(\boldsymbol{\eta}) - \exp(\boldsymbol{\eta}^0) \right)^\top \mathbf{G}_{gg} \left(\exp(\boldsymbol{\eta}) - \exp(\boldsymbol{\eta}^0) \right), \\ \gamma_{i,n}(\boldsymbol{\eta}) &= \mathbf{g}_i^\top \left(\exp(\boldsymbol{\eta}) - \exp(\boldsymbol{\eta}^0) \right) / R_n(\boldsymbol{\eta}), \end{aligned}$$

from which it follows that

$$R_n(\boldsymbol{\eta}) \left| \sum_{i=1}^n \gamma_{i,n}(\boldsymbol{\eta}) \epsilon_i \right| = \left| \sum_{i=1}^n \left(\exp(\boldsymbol{\eta}) - \exp(\boldsymbol{\eta}^0) \right)^\top \mathbf{g}_i \epsilon_i \right| \leq n \sum_{j=1}^m \left| \exp(\eta_j) - \exp(\eta_j^0) \right| \left| n^{-1} \sum_{i=1}^n \mathbf{g}_i \epsilon_i \right|.$$

Under assumption 2 and given that, by definition, ϵ_i are i.i.d. disturbances with finite variance σ^2 , $n^{-1} \sum_{i=1}^n \mathbf{g}_i \epsilon_i$ tends to zero in mean squared error and so in probability as well. Indicate with Ψ the set of all possible state price density and notice that $R_n(\boldsymbol{\eta}) \geq 0$ due to the definition of the model matrix. Then, following Malinvaud (1970, section 3), in order to prove that $\exp(\boldsymbol{\eta})$ is a consistent estimator of $\exp(\boldsymbol{\eta}^0)$, it is sufficient to show that for any closed $\mathcal{H} \subset \Psi$ not containing $\exp(\boldsymbol{\eta}^0)$

$$\sup_{\exp(\boldsymbol{\eta}) \in \mathcal{H}} \left| \exp(\boldsymbol{\eta}) - \exp(\boldsymbol{\eta}^0) \right| R_n^{-1} \quad (\text{A.1})$$

is bounded by a quantity not depending on n .

Indicate with d the Euclidean distance between $\exp(\boldsymbol{\eta})$ and $\exp(\boldsymbol{\eta}^0)$ and with ν_n the smallest eigenvalue of \mathbf{G}_{gg} . Assumption 3 implies that ν_n tends to the smallest eigenvalue of $\bar{\mathbf{G}}$, say $\bar{\nu} > 0$ and, for n sufficiently large, $\nu_n \geq 0.5\bar{\nu}$ is verified. Noticing that $\sqrt{d} \geq |\exp(\boldsymbol{\eta}) - \exp(\boldsymbol{\eta}^0)|$, it follows that

$$n \left| \exp(\boldsymbol{\eta}) - \exp(\boldsymbol{\eta}^0) \right| R_n^{-1}(\boldsymbol{\eta}) < 2 \frac{\sqrt{d}}{\bar{\nu}},$$

since $n^{-1} R_n(\boldsymbol{\eta}) \geq \nu_n d$. This guarantees that, for any bounded $\mathcal{H} \subset \Psi$ not containing $\exp(\boldsymbol{\eta}^0)$, \sqrt{d} is bounded below by a positive quantity and hence Eq. (A.1) is bounded by a

quantity not depending on n . This result completes our proof. Indeed, given that the inverse mapping from $\exp(\mathcal{E})$ to \mathcal{E} is one-to-one and continuous for all $\boldsymbol{\eta} \in \mathcal{E}$ (and hence also at $\boldsymbol{\eta}^0$), we can assert that $\boldsymbol{\eta}$ is a consistent estimator of $\boldsymbol{\eta}^0$.

Proof of Proposition 2

The existence of the series of estimators $\boldsymbol{\eta}$ for $\lambda = o(n^{-1/2})$ can be proved following Yu and Ruppert (2002, appendix B). We now want to prove the asymptotic normality of these estimators.

For consistent estimators $\boldsymbol{\eta}$, a linearization of $S_n(\boldsymbol{\eta})$ around $\boldsymbol{\eta}^0$ leads to

$$0 = \left. \frac{\partial S_n}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}} \approx \left. \frac{\partial S_n}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}^0} + \left. \frac{\partial^2 S_n}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \right|_{\boldsymbol{\eta}^0} (\boldsymbol{\eta} - \boldsymbol{\eta}^0),$$

where $\partial^2 S_n / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top$ always exists since $\mu(\boldsymbol{\eta})$ is two times continuously differentiable by definition. From the equation above we can write

$$\sqrt{n}(\boldsymbol{\eta} - \boldsymbol{\eta}^0) = - \left[n^{-1} \frac{\partial^2 S_n}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \right]^{-1} n^{-1/2} \frac{\partial S_n}{\partial \boldsymbol{\eta}}.$$

The asymptotic normality of the proposed estimators can be verified following Amemiya (1983, par. 2.2.2) and requires that

$$n^{-1/2} \left. \frac{\partial S_n}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}^0} \rightarrow \mathcal{N}(0, 4\sigma^2 \mathcal{K}(\boldsymbol{\eta}^0)),$$

and

$$\frac{\partial^2 S_n}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \rightarrow 2\mathcal{K}(\boldsymbol{\eta}^0) \text{ as } n \rightarrow \infty.$$

By differentiating $S_n(\boldsymbol{\eta})$ w.r.t. $\boldsymbol{\eta} = \boldsymbol{\eta}^0$ we obtain

$$\begin{aligned} n^{-1/2} \left. \frac{\partial S_n}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}^0} &= 2n^{-1/2} \sum_{i=1}^n (c_i - \mu_i(\boldsymbol{\eta})) \left. \frac{\partial \mu(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}^0} + 2n^{1/2} \lambda_n \mathbf{D}^\top \mathbf{D} \boldsymbol{\eta}^0 \\ &= -2n^{-1/2} \sum_{i=1}^n \epsilon_i \left. \frac{\partial \mu(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}^0} + 2n^{1/2} \lambda_n \mathbf{D}^\top \mathbf{D} \boldsymbol{\eta}^0. \end{aligned}$$

The term $\sum_{i=1}^n \epsilon_i \left. \frac{\partial \mu(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}^0}$ is a weighted sum of i.i.d. (by hypothesis) error terms and so it converges in probability to $\mathcal{N}(0, 4\sigma^2 \mathcal{K}(\boldsymbol{\eta}^0))$. This result is obtained by central limit theorem taking into account assumption 3 and recalling that $2n^{1/2} \lambda_n \mathbf{D}^\top \mathbf{D} \boldsymbol{\eta}^0$ tends to zero since

$\lambda = o(n^{-1/2})$ by assumption.

In order to prove that $\partial^2 S_n / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top \rightarrow 2\boldsymbol{\mathcal{K}}(\boldsymbol{\eta}^0)$ as $n \rightarrow \infty$ we need to compute the second derivative of $S_n(\boldsymbol{\eta})$ for $\boldsymbol{\eta}$ in a neighborhood of $\boldsymbol{\eta}^0$

$$\begin{aligned} n^{-1} \frac{\partial^2 S_n}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} &= n^{-1} \frac{\partial^2 \|\mathbf{c} - \boldsymbol{\mu}(\boldsymbol{\eta})\|^2}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} + 2n^{1/2} \lambda_n \mathbf{D}^\top \mathbf{D} \\ &= \frac{2}{n} \sum_{i=1}^n \frac{\partial \mu_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\partial \mu_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^\top} - \frac{2}{n} \sum_{i=1}^n (\mu_i(\boldsymbol{\eta}^0) - \mu_i(\boldsymbol{\eta})) \frac{\partial^2 \mu_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \\ &\quad - \frac{2}{n} \sum_{i=1}^n \epsilon_i \frac{\partial^2 \mu_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} + 2n^{1/2} \lambda_n \mathbf{D}^\top \mathbf{D}. \end{aligned}$$

Since $\lambda_n = o(n^{-1/2})$ the last term of the second derivative goes to zero and, given the consistency of the DESPD estimator and under assumption 3, it follows that $\frac{\partial^2 S_n}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top}$ converges in probability to $2\boldsymbol{\mathcal{K}}(\boldsymbol{\eta}^0)$ for $\boldsymbol{\eta}$ in a neighborhood of $\boldsymbol{\eta}^0$. Then, by using Slutsky's theorem

$$\sqrt{n}(\boldsymbol{\eta} - \boldsymbol{\eta}^0) \rightarrow \mathcal{N}(0, \sigma^2 \boldsymbol{\mathcal{K}}^{-1}(\boldsymbol{\eta}^0))$$

is verified.